

AAAAATAAATCAT ATG AAA AAT ATT AAA AAT CAA GTA ATC AAT CTC GGT CCT AAT TCT  
 K M I K K N I K N Q V M N L G P N S  
 AAA TTA TTA AAA GAA TAT AAA TCA CAA TTA ATT GAA TTA AAT ATT GAA CAA TTT GAA GCA  
 K L L K E Y K S Q L I I E L N I I E Q F E A  
 GGT ATT GGT TTA ATT TTA GGA GAT GCT TAT ATT CGT AGT CGT GAT GAA GGT AAA ACT TAT  
 G I G L I L G D A Y I I R S R D E G K T Y  
 TGT ATG CAA TTT CAC TCC AAA AAT AAG GCA TAC ATG GAT CAT GTA TGT TTA TTA TAT GAT  
 C M Q F E W K N K A Y M D H V C L L Y D  
 CAA TGG GTA TTA TCA CCT CCT CAT AAA AAG GAA AGA GTT AAT CAT TTA GGT AAT TTA GTA  
 Q W V L S P P H K K E R V N H L G N L V  
 ATT ACC TGG GGA GCT CAA ACT TTT AAA CAT CAA GCT TTT AAT AAA TTA GCT AAC TTA TTT  
 I T W G A Q T F K K H Q A F N K L A N L F  
 ATT GTA AAT AAT AAA CTT ATT CTT AAT AAT TTA GTT GAA AAT TAT TTA ACA CCT ATG  
 I Y N N K K L I I P N N L V E N Y L T P M  
 AGT CTG GCA TAT TGG TTT ATG GAT GAT GGA GGT AAA TGG GAT TAT AAT AAA AAT TCT CTT  
 S L A Y W F M D D G G K W D Y N K N S L  
 AAT AAA AGT ATT GTA TTA AAT ACA CAA AGT TTT ACT TTT GAA GAA GTA GAA TAT TTA CTT  
 N K S I V L N T Q S F T F E E V C Y L V  
 AAA GGT TTA AGA AAT AAA TTT CAA TTA AAT TGT TAT GTT AAA ATT AAT AAA AAT AAA CCA  
 K G L R N K F Q L N C Y V K I I N K N K P  
 ATT ATT TAT ATT GAT TCT AGT AGT TAT CTG ATT TTT TAT AAT TTA ATT AAA CCT TAT TTA  
 I I I Y I D S M S Y L I F Y N I I T K P Y L  
 ATT CCT CAA ATG ATG M Y K L P N T I S TCC GAA ACT TTT TTA AAA TAA  
 I P Q M Y K L P N T I S TCC GAA ACT TTT TTA AAA TAA

FIG. 1

Bam HI

1. CCGGATCCATG CAT ATG AAA AAC ATC AAA AAC CAG GTA ATG AAC CTG GGT CCG AAC TCT  
M H M K N I K N Q V M N L G P N S  
AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC ATC GAA CAG TTC GAA GCA  
K L L K E Y K S Q L I I E L N I I E Q F E A  
GGT ACT GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT GAT GAA GGT AAA ACC TAC  
G I G L I L G D A Y I I R S S R D E G K T Y  
TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC GTA TGT CTG CTG TAC GAT  
C M Q F E W K N K A A Y M D H V C L L Y D  
CAG TGG GTA CTG TCC CCG CCG CAG ACT TTC AAA CAC CAA CGT GTT AAC CAC CTG GGT AAC CTG GTA  
Q W V L L S P P H K K E R V N H L G N L V  
ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC AAA CTG GCT AAC CTG TTC  
I T W G A Q T F K H Q A F N K L A N L F  
ATC GTT AAC AAC AAA ACC ATC CCG AAC AAC CTG GTT GAA AAC TAC CTG ACC CCG ATG  
I V N N K K T I P N N L V E N Y L T P M  
TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GAT TAC AAC AAA AAC TCT ACC  
S L A Y W F M D D G W D Y N K N S T  
AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA GAA GTA TAC CTG GTT  
N K S I V L N T Q S F T F E E V E Y L V  
AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC CTA AAA ATC AAC AAA AAC CCG  
K G L R N K F Q L N C Y Y K I N K N K P  
ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC CTG ATC AAA CCG TAC CTG  
I I Y I D S M S Y L I F Y N L I K P Y L  
ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC GAA ACT TTC CTG AAA TAA  
I P Q M M Y K L P N T I S S E T F L K  
TAAGTCGACTGCAGGATCCGGTAAGTAAGTAA  
I I I I I  
Sall PstI BamHI
- 2.

1 and 2: THESE AMINO ACIDS ARE ABSOLUTELY NECESSARY TO PRODUCE CATALYTIC ACTIVITY. OTHER SUBSTITUTIONS ARE POSSIBLE, SUCH AS DELETIONS OF THE 10 FIRST AMINO ACIDS. FIG. 2

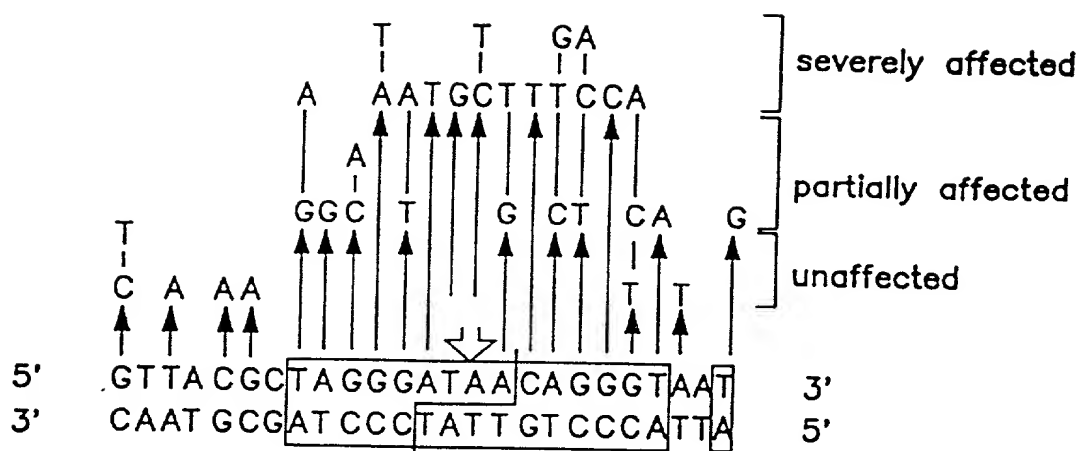


FIG. 3

1667 GCGGACAGGTATCCGGTAAGCGGAGGGTCGGAACAGGAGAGCGACGAGGGAGCTTCCAGGGGGAACGCGCTGGTATCT 1746  
1747 TTATAGTCCIGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTG ATG CTC GTC AGG GGG GCG GAG 1818  
1819 CCT ATG GAA AAA CGC CAG CAA CGC GGC CTT TTT ACG GTT CCT GGC CTT TTG CTG GCC TTT 1878  
1879 TGC TCA CAT GTT CTT TCC TGC GTT ATC CCC TGA TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC 1947  
1948 TGATACCGCTCGCCGCGAGCCGAACGAGCGAGCGAGTCAGTGAGCGGAGGAGCGGCCCAATACGCAAAC 2027  
2028 CGCCTCTCCCCGCGGTTGGCCGATTTCATTA ATG CAG CTG GCA CGA CAG GTT TCC CGA CTG GAA AGC 2094  
2095 GGG CAG TGA GCGCAACGCAATTA ATG TGA GTTAGCTCACTCATTAGGCACCCCGAGGCTTTACACTTT ATG 2164  
2165 CTT CCG GCT CGT ATG TTG TGT GGA ATT GTG AGC GGA TAA CAATTTACACAGGAACAGCT ATG 2228  
2229 ACC ATG ATT ACG AAT TCT CAT GTT TGA CAGCTTATCATCGATAAGCTTTA ATG CCG TAG TTTATCAC 2295  
2296 AGTTAAATTGCTAACGCGAGTCAGGCACCGTGT ATG AAA TCT AAC AAT GCG CTC ATC GTC ATC CTC GGC 2363  
2364 ACC GTC ACC CTG GAT GCT GTA GGC ATA GGC TTG GTT ATG CCG GTA CTG CCG GGC CTC TTG 2423  
2424 CGG GAT ATC CGC CTG ATG CGT GAA CGT GAC GGA CGT AAC CAC CGC GAC ATG TGT GTG CTG 2483  
2484 TTC CGC TGG GCA TGC CAG GAC AAC TTC TGG TCC GGT AAC GTG CTG AGC CCG GCC AAG CTT 2543

FIG. 4A

2544	ACT	CCC	CAT	CCC	CCT	GTT	GAC	AAT	TAA	TCATCGGCTCGTATA	ATG	TGT	GGA	ATT	GTG	AGC	GGA	2606
73	T	P	H	P	P	V	D	N	*			C	G	I	V	S	G	7
2607	TAA	CAATTT	CACACAGGAACAGGATCC															
8	*																	
2671	AAC	CTG	GGT	CCG	AAC	TCT	AAA	CTG	CTG	AAA	AAC	ATC	AAA	AAA	AAC	CAG	GTA	ATG
13	N	L	G	P	P	N	S	K	L	L	K	L	K	Y	K	N	Q	M
2730																		12
2731	ATC	GAA	CAG	TTC	GAA	GCA	GGT	ATC	GGT	CTG	ATC	CTG	GGT	GAT	TAC	ATC	CGT	2730
33	I	E	Q	F	E	A	G	I	G	L	I	L	G	D	A	Y	I	32
2790																		52
2791	GAT	GAA	GGT	AAA	ACC	TAC	TGT	ATG	CAG	TTC	GAG	TGG	AAA	AAC	GCA	TAC	ATG	2790
53	D	E	G	K	T	Y	C	M	Q	F	E	W	K	N	K	A	Y	52
2850																		72
2851	GTA	TGT	CTG	CTG	TAC	GAT	CAG	TGG	GTA	CTG	TCC	CCG	CAC	AAA	AAA	GAA	CGT	2850
73	V	C	L	L	Y	D	Q	W	V	L	S	P	H	K	K	E	R	72
2910																		92
2911	CAC	CTG	GGT	AAC	CTG	GTA	ATC	ACC	TGG	GGC	GCC	CAG	ACT	TTC	AAA	CAC	CAA	2910
93	H	L	G	N	L	V	I	T	W	G	A	Q	T	F	K	H	Q	112
2970																		112
2971	AAA	CTG	GCT	AAC	CTG	TTC	ATC	GTT	AAC	AAC	AAA	ACC	ATC	CCG	AAC	AAC	CTG	2970
113	K	L	A	N	L	F	I	V	N	N	K	K	T	I	P	N	N	132
3030																		132
3031	AAC	TAC	CTG	ACC	CCG	ATG	TCT	CTG	GCA	TAC	TGG	TTC	ATG	GAT	GGT	GGT	AAA	3030
133	N	Y	L	T	P	M	S	L	A	Y	W	F	M	D	G	G	K	152
3090																		152
3091	TAC	AAC	AAA	AAC	TCT	ACC	AAC	AAA	TCG	ATC	GTA	CTG	AAC	ACC	CAG	TCT	TTC	3090
153	Y	N	K	N	S	T	N	K	S	I	V	L	N	T	Q	S	F	172
3150																		172
3151	GAA	GTA	GAA	TAC	CTG	GTT	AAG	GGT	CTG	CGT	AAC	AAA	TTC	CAA	CTG	AAC	TGT	3150
173	E	V	E	Y	L	V	K	G	L	R	N	K	F	Q	L	N	C	192
3210																		192
3211	ATC	AAC	AAA	AAC	AAA	CCG	ATC	ATC	TAC	ATC	GAT	TCT	ATG	TCT	TAC	CTG	ATC	3210
193	I	N	K	N	K	P	I	I	Y	I	D	S	M	S	Y	L	I	212
3270																		212
3271	CTG	ATC	AAA	CCG	TAC	CTG	ATC	CCG	CAG	ATG	ATG	TAC	AAA	CTG	CCG	AAC	ACT	3270
213	L	I	K	P	Y	L	I	Q	M	M	M	Y	K	L	P	N	I	232
3330																		232
3331	GAA	ACT	TTC	CTG	AAA	TAA												3404
233	E	T	F	L	K	*												238

SalI PstI

FIG. 4B

	-2	-1	1				5					10						
	M	H	M	K	N	I	K	K	N	Q	V	M	N	L	G	P	N	S
K	L	L	20	K	E	Y	K	S	Q	L	I	E	L	30	N	I	E	Q
G	I	G	40	L	I	L	G	D	A	Y	I	R	S	50	R	D	E	G
C	M	Q	60	F	E	W	K	N	K	A	Y	M	D	70	H	V	C	L
Q	W	Y	80	L	S	P	P	H	K	K	E	R	Y	90	N	H	L	G
I	T	W	100	G	A	Q	T	F	K	H	Q	A	F	110	N	K	L	A
I	V	N	120	N	K	K	I	I	P	N	N	L	V	130	E	N	Y	L
G	L	A	140	Y	W	P	M	D	D	G	G	K	W	150	D	Y	N	K
N	K	S	160	I	V	L	N	T	Q	S	F	T	F	170	E	E	V	E
K	G	L	180	R	N	K	F	Q	L	N	C	Y	V	190	K	I	N	K
I	I	Y	200	I	D	S	M	S	Y	L	I	F	Y	210	N	L	I	K
I	P	Q	220	M	Y	K	L	P	N	T	I	S	230	S	E	T	F	L

Positions that can be changed without affecting enzyme activity (demonstrated)  
positions -1 and -2 are not natural. The two amino acids are added due to cloning strategies

positions 1 to 10: can be deleted  
position 36: G is tolerated  
position 40: M or V are tolerated  
position 41: S or N are tolerated  
position 43: A is tolerated  
position 46: V or N are tolerated  
position 91: A is tolerated  
positions 123 and 156: L are tolerated  
position 223: A and S are tolerated

#### Changes that affect enzyme activity (demonstrated)

position 19: L to S  
position 38: I to S or N  
position 39: G to D or R  
position 40: L to Q  
position 42: L to R  
position 44: D to E, G or H  
position 45: A to E or D  
position 46: Y to D  
position 47: I to R or N  
position 80: L to S  
position 144: D to E  
position 145: D to E  
position 146: G to E  
position 147: G to S

FIG. 5

Group I Intron Encoded Endonucleases and Related Endonucleases

ENDONUCLEASE	RECOGNITION SEQUENCE	CLEAVAGE SITE	▽ INTRON SITE
TWO DODECAPEPTIDE FAMILY (OR 4 BP CUTTERS)	I-Sce I (Saccharomyces mitochondria)	CGCTAGGGATAA CAGGGTAATATAGC GCGATCCCTATTGTCCCATTAATATCG	
	I-Sce IV (Saccharomyces mitochondria)	TTCTCATGATTAGCTCTAATCCATGG AAGAGTACTAATCGAGATTAGGTACC	
	I-Sce II (Saccharomyces mitochondria)	CTTTGGTCA CCAGAAAGTATATATTT GAAACCAGTAGGTCTTCATATATAAAA	
	I-Ceu I (Chlamydomonas chloroplast)	TAA CGGTCTTAA AGGTAGCGAAATTCA ATTGCCAGGATTCCATCGCTTTAAGT	
	I-Ppo I (Physarum nucleus)	TGACTCTCTTAA AGGTAGCCAAATGCC ACTGAGAGAATTCCATCGGTTTACGG	
	I-Sce III (Saccharomyces mitochondria)	GGAGGTTTTGGTAACTATTTATTACC CCTCCAAAACCATTGATAAATAATGG	
	I-Cre I (Chlamydomonas chloroplast)	GGGTTCAAAACGTCTGTGAGACAGTTT CCCAAGTTTTGCAGCACTCTGTCAA	
	Endo. Sce I(RF3) (Saccharomyces mitochondria) (Non intronic)	GATGCTGTAGGCATAGGCTTGGTTAT CTACGACATCCGTATCCGAACCAATA	
	HO (Saccharomyces nucleus) (Non intronic)	CTTTCCGCAACA GTATAATTTTATAA GAAAGGCGTTGTCAATTATTAATAATT	
	I-Csm I (Chlamydomonas mitochondria) (Putative endonuclease)	ACCATGGGGTCAAATGTCTTTCTGGG TGGTACCCAGTTTACAGAAAGACCC	
OTHER STRUCTURAL FAMILIES	I-Pan I (Podospira mitochondria) (Putative endonuclease)	GTGCCTGAATGATATTTATTACCTTT CACGGACTTACTATAAATAATGAAA	
	(Bacteriophage T4)	I Tev I CAAC GCTCAGTAGATGTTTTCTTGGGTCTACCGTTTAAAT GTTGCGAGTCATCTACAAAAGAA CCAGATGGCAAATTA	
	I Tev II	CAAGCTTATGAGTATGAAGTGAACACGTTATT GTTCGAATACTCATACTTCACTTGTG CAATAA	
	I Tev III	GCTATTGTTTTT ATGTATCTTTTGCCTGTAGCTTTAA CGATAAGCAAAAAT ACATAGAAAACGCACATCGAAATT	

FIG. 6

# EXPRESSION VECTORS

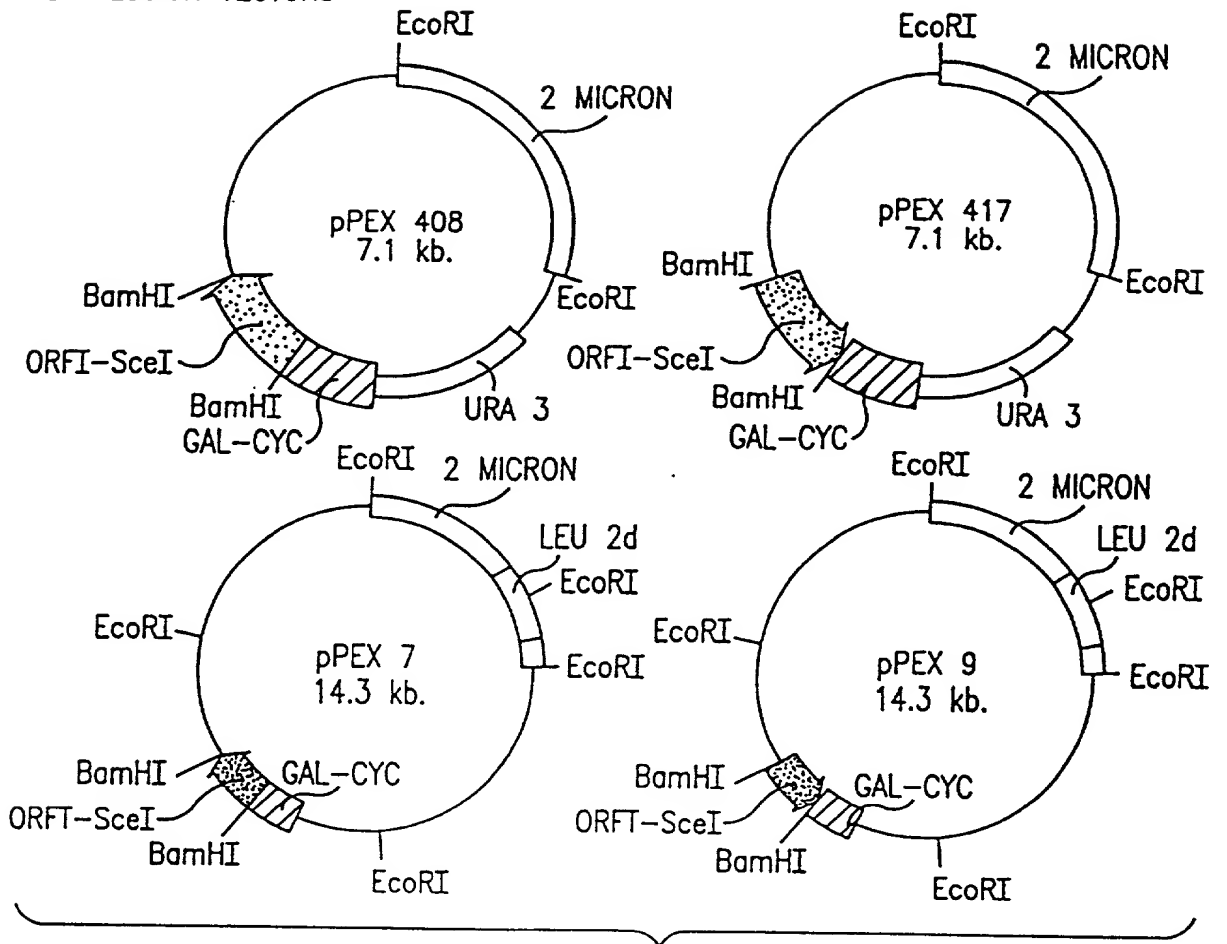


FIG. 7

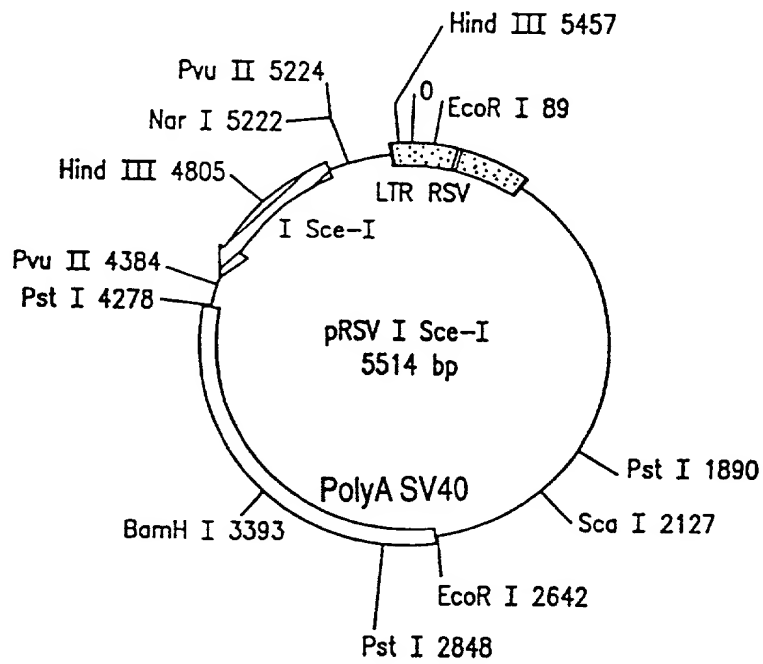
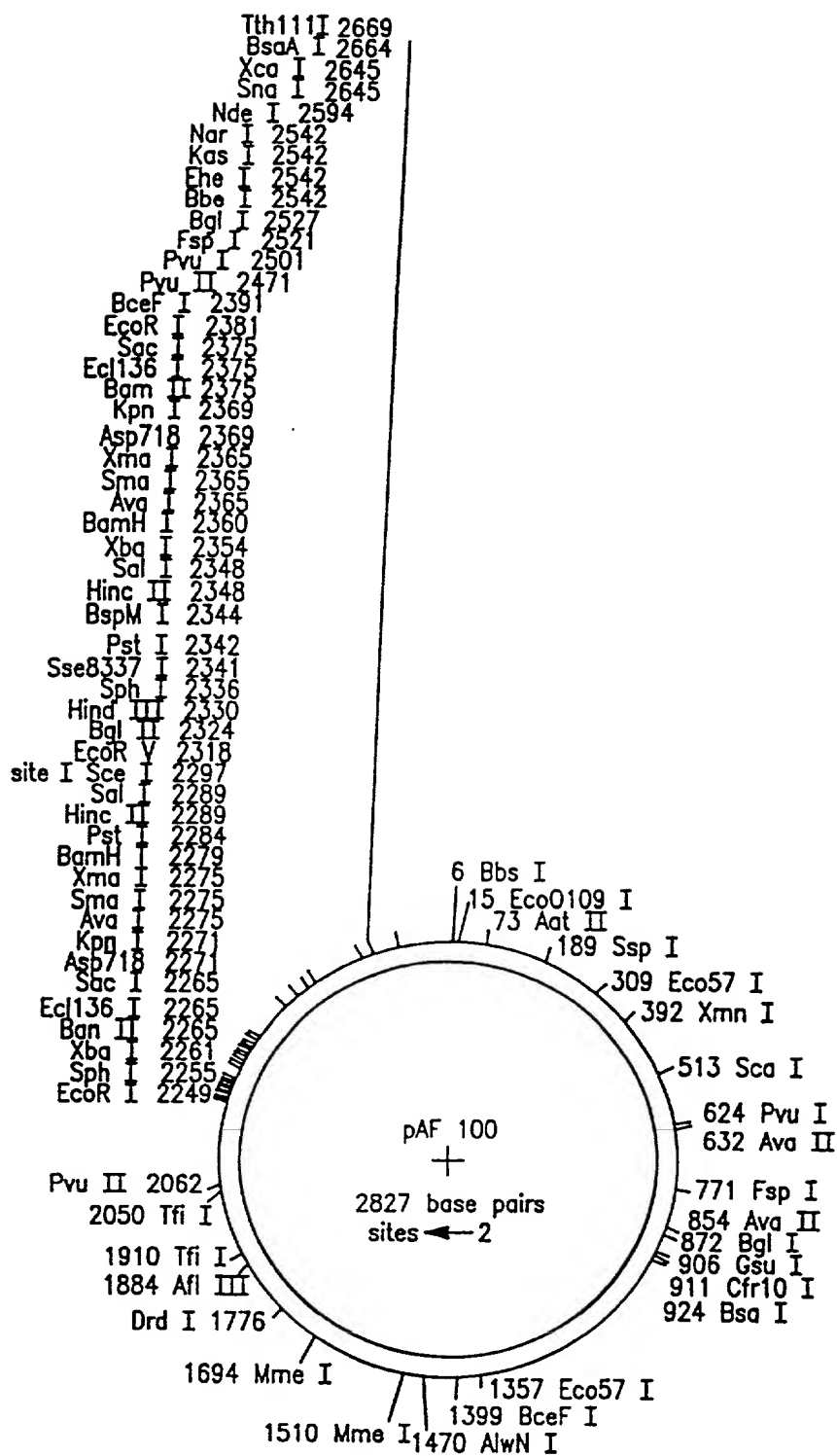


FIG. 8





00036169-041001

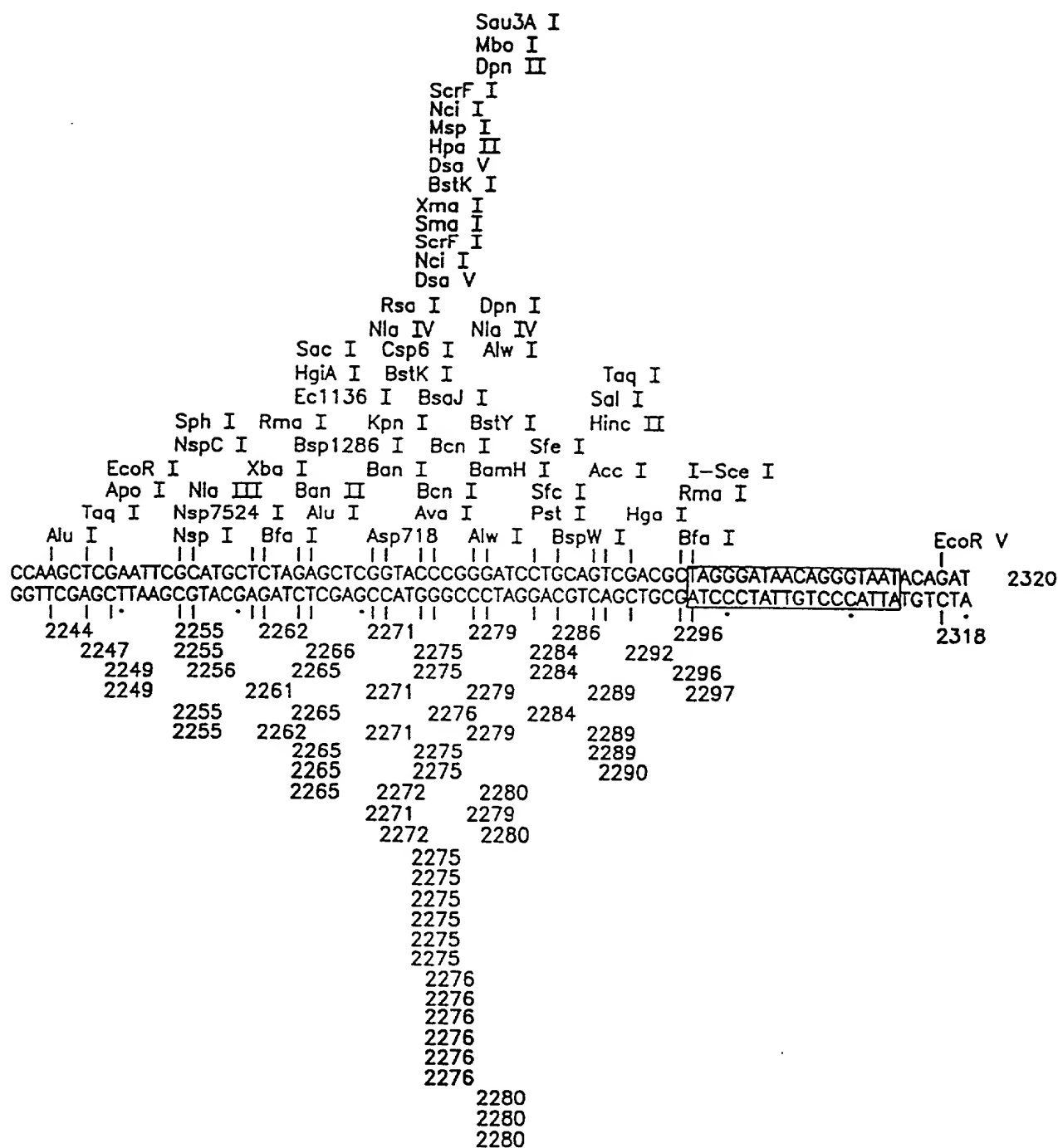


FIG. IOA

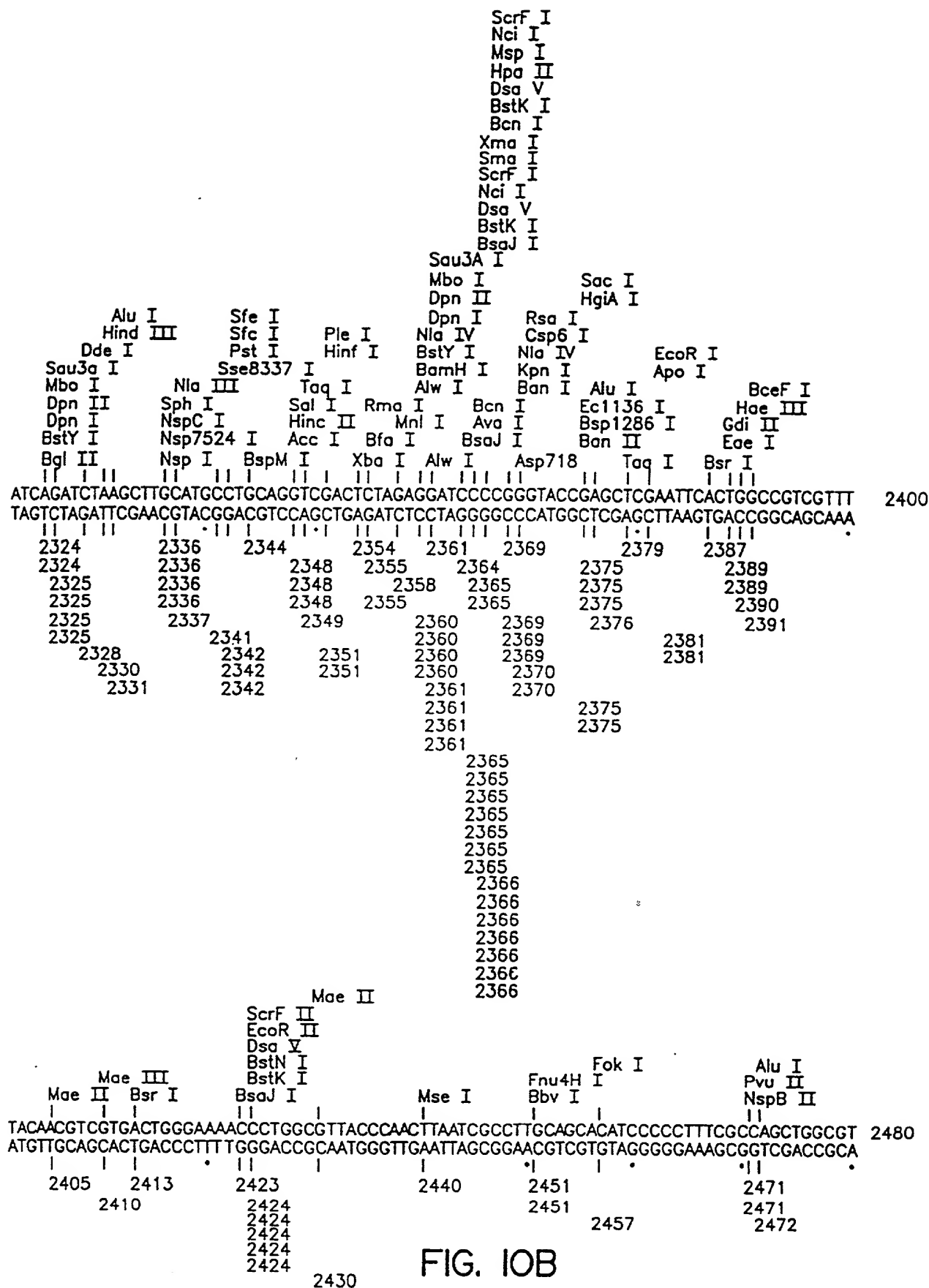
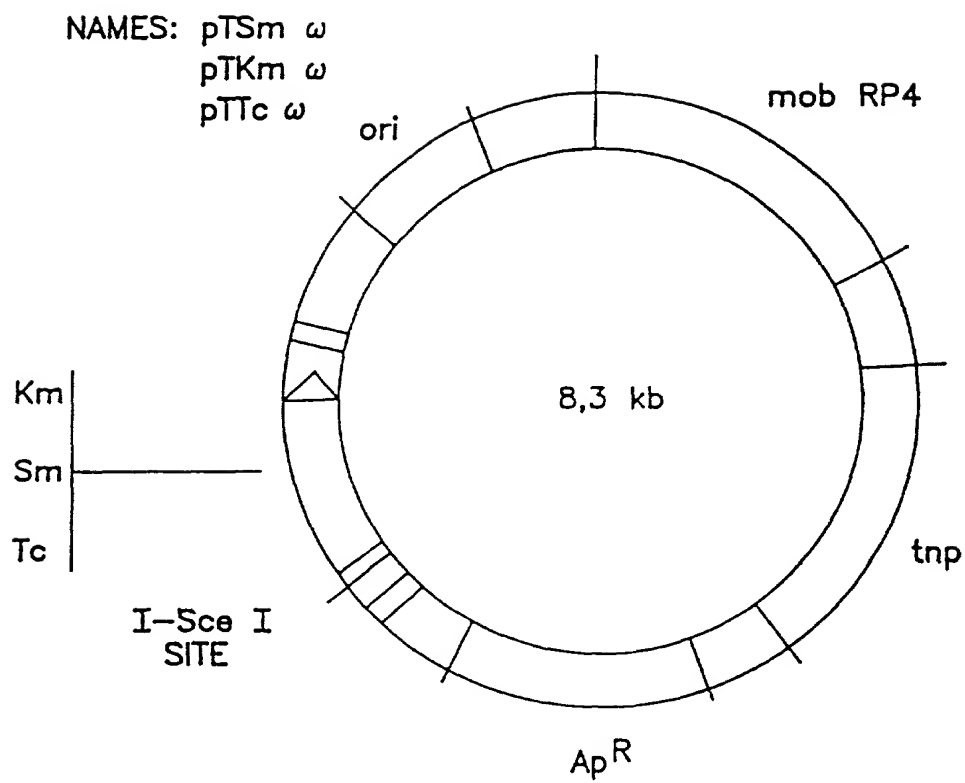
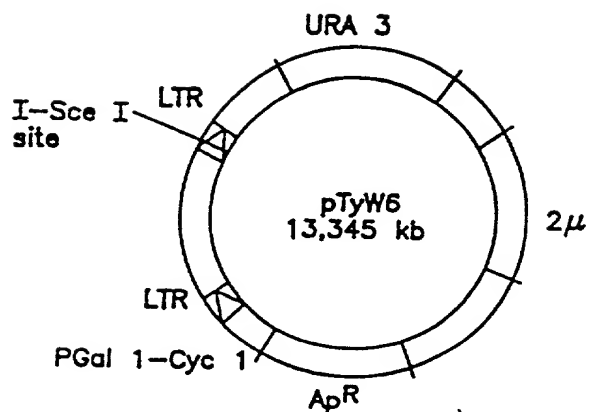


FIG. 10B



Construction: pGP 704 from De Lorenzo, with transposase  
 gene and insertion of the linker[I-SceI] in NotI unique site

FIG. 11



Construction: pD 123, from J.D. Boeke with insertion of a linker[I-SceI-NotI] in BamHI

FIG. 12

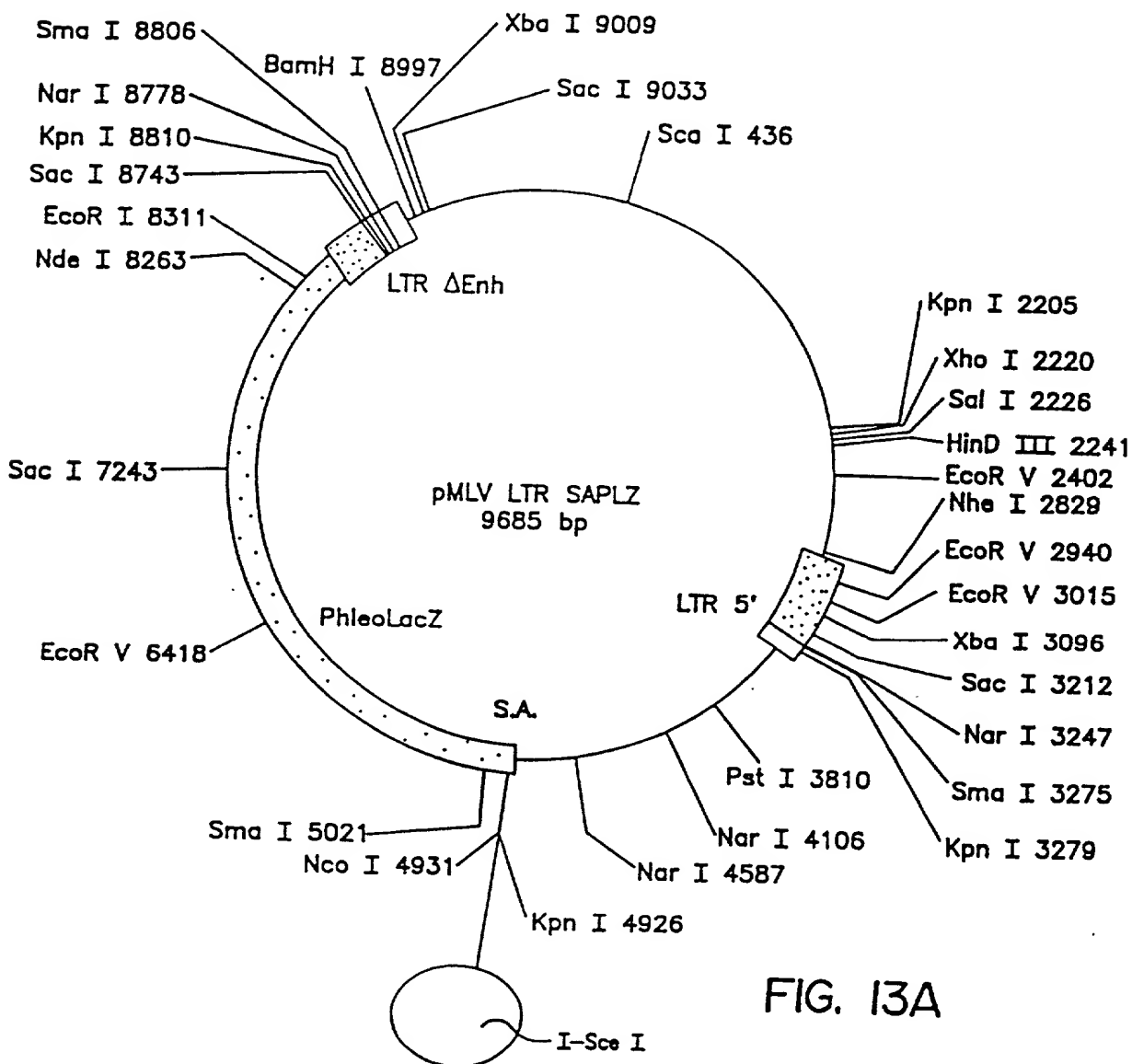


FIG. 13A

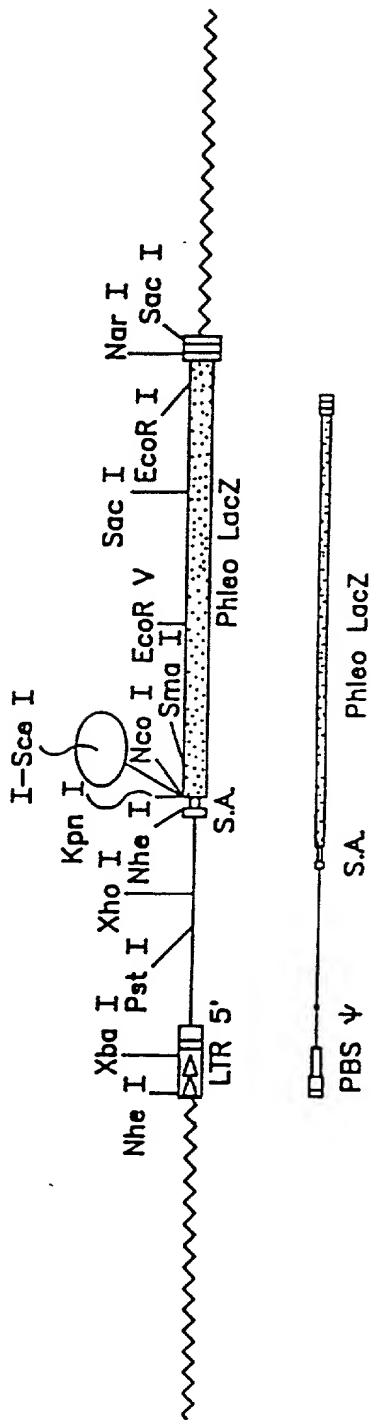


FIG. 13B

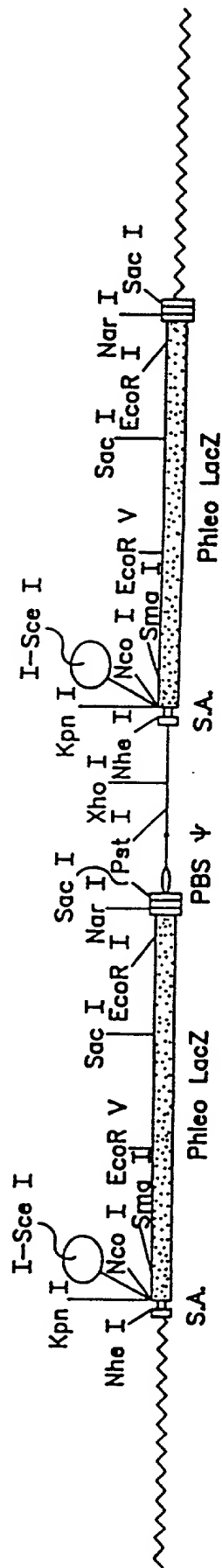
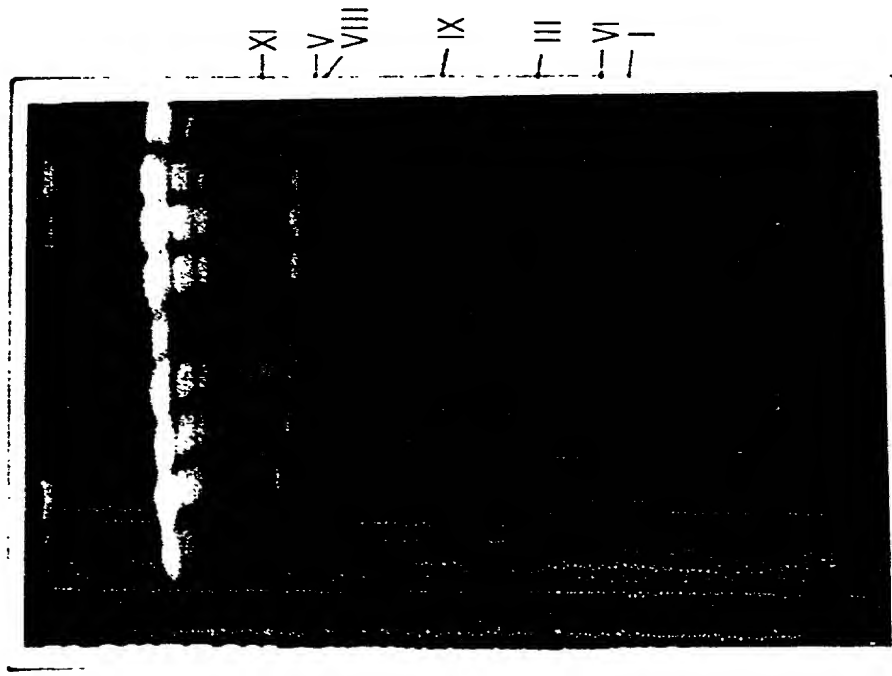
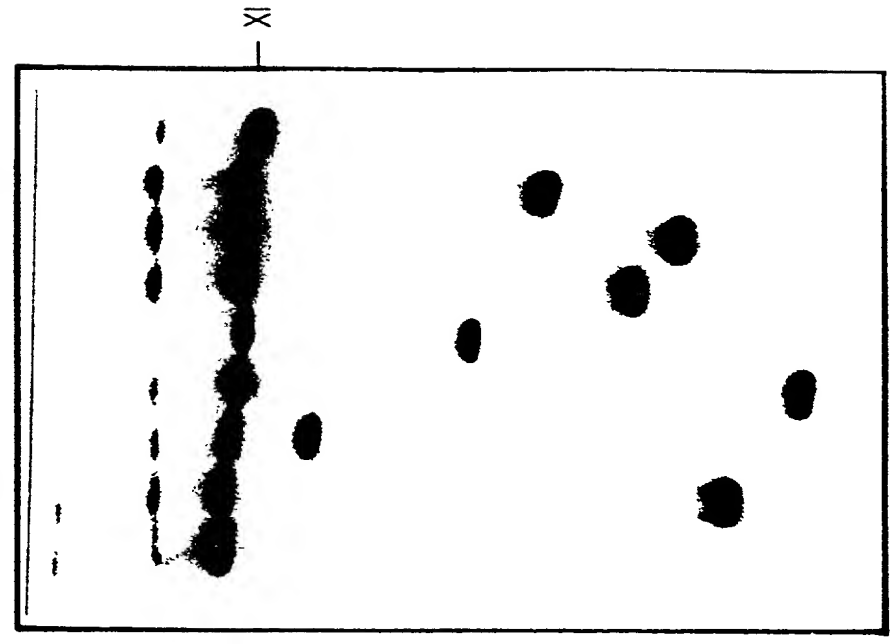


FIG. 13C

CONTROL D304 A302 CONTROL A304 E40 G41 M57 H81 T62 CONTROL



CONTROL A302 A304 E40 G41 M57 H81 T62 CONTROL



LEFT END PROBE  
COSMID pUKG 040

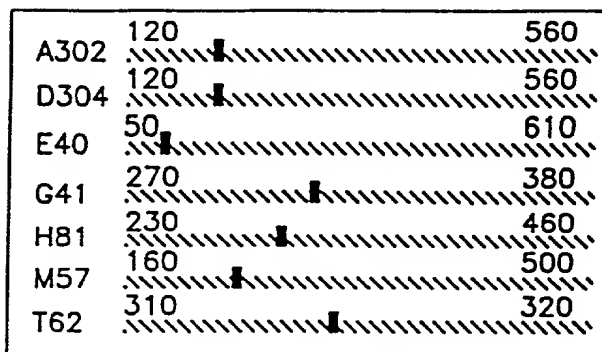


FIG. 15A

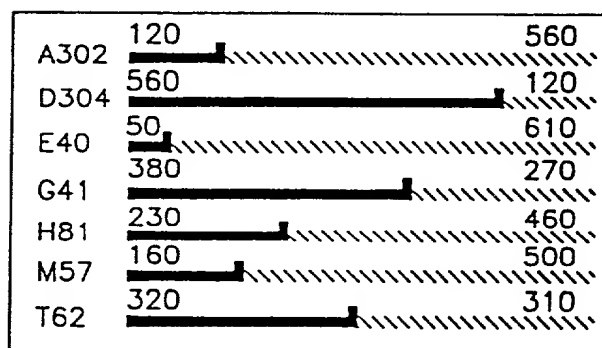


FIG. 15B

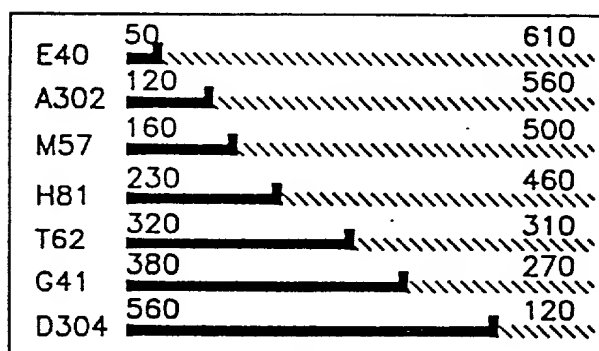
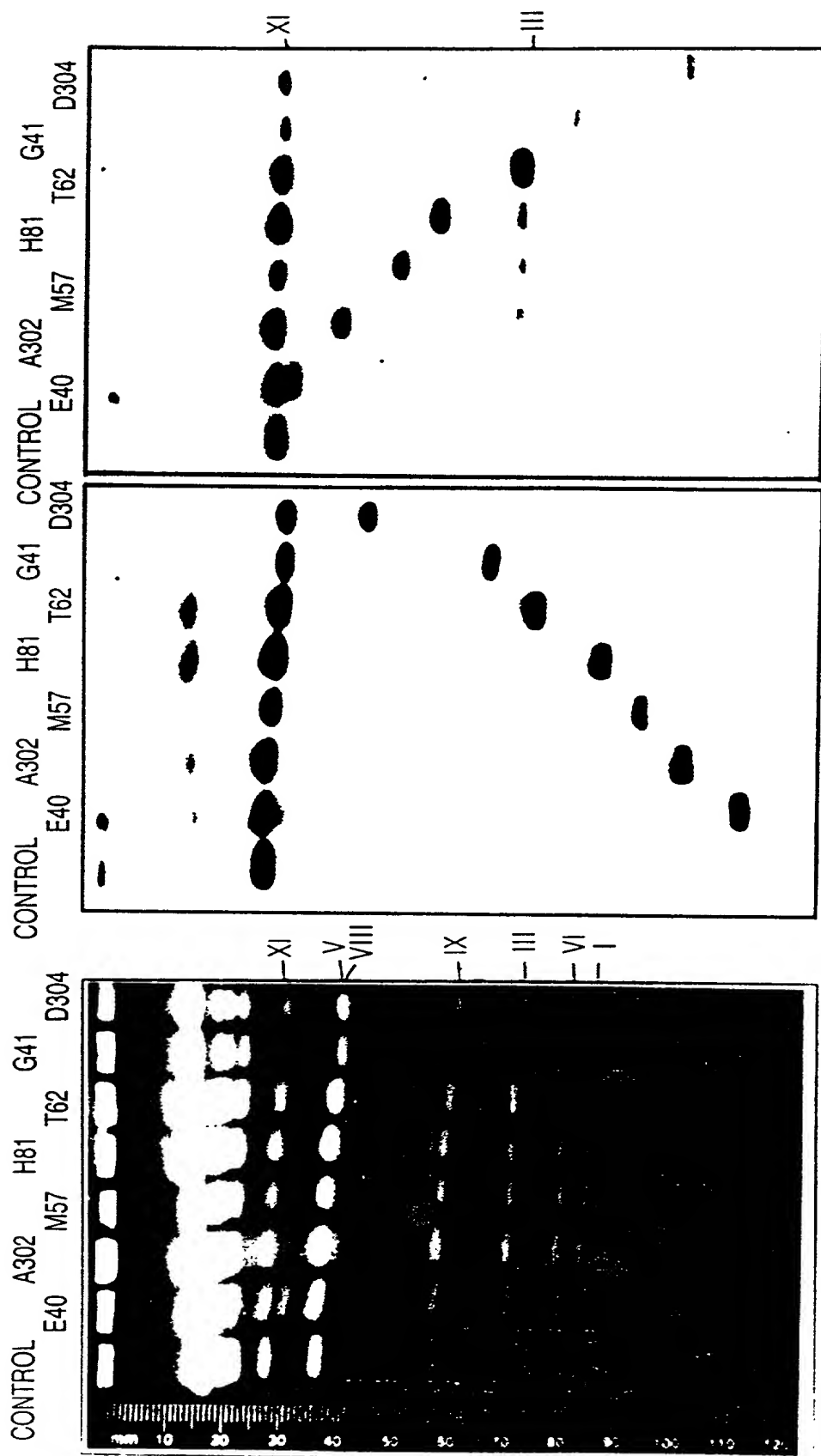


FIG. 15C



FIG. 15D



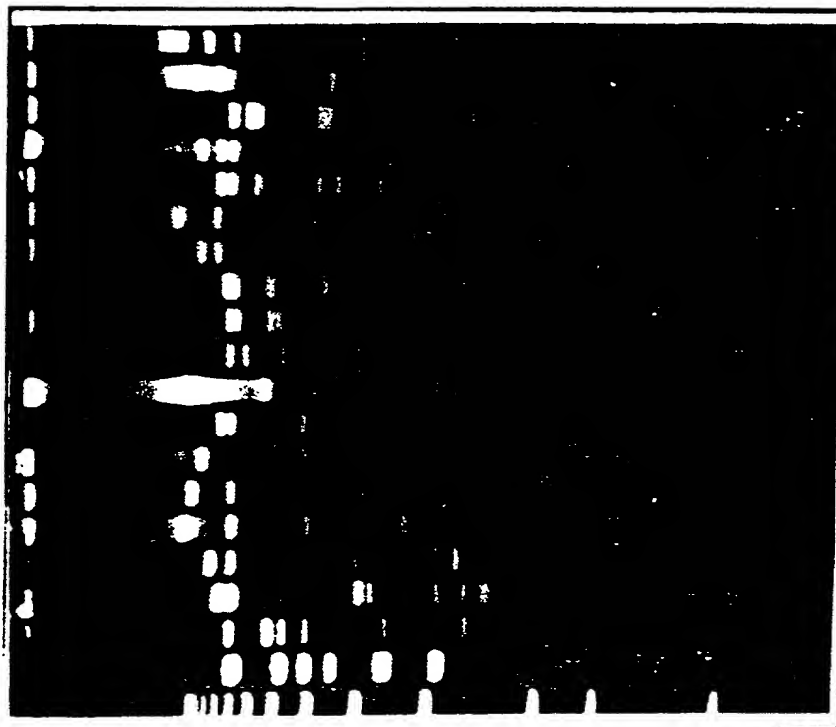
RIGHT END PROBE  
COSMID pUKG066

LEFT END PROBE  
COSMID pUKG040

FIG. 16A

FIG. 16B

FIG. 16C



pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG  
019 097 081 121 119 021 146 144 046

pEKG pEKG pEKG pUJK pUJK pEKG pEKG pEKG pUJK  
100 098 090 151 148 011 047 013 118 066

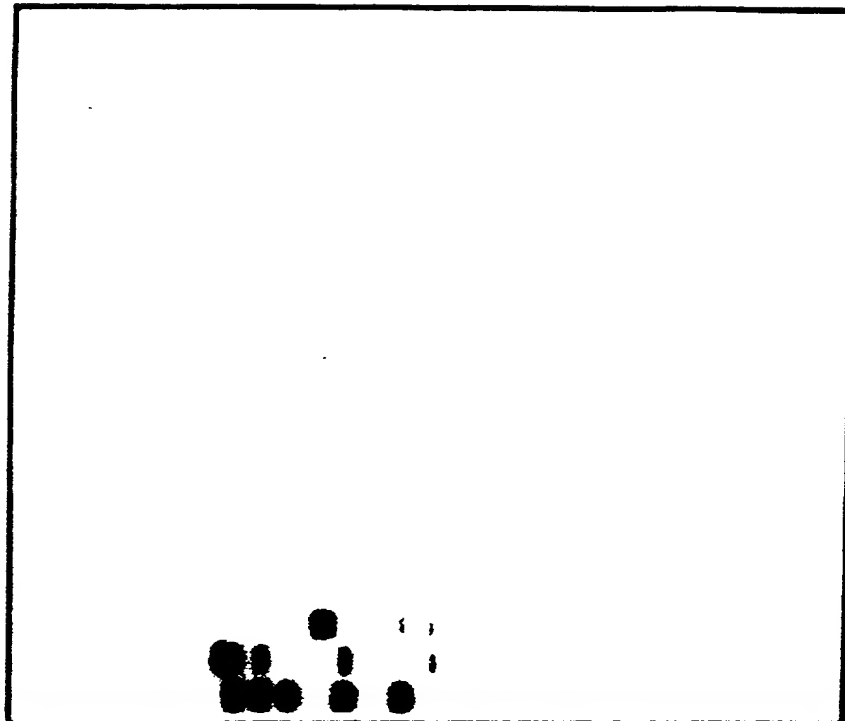


FIG. 17B

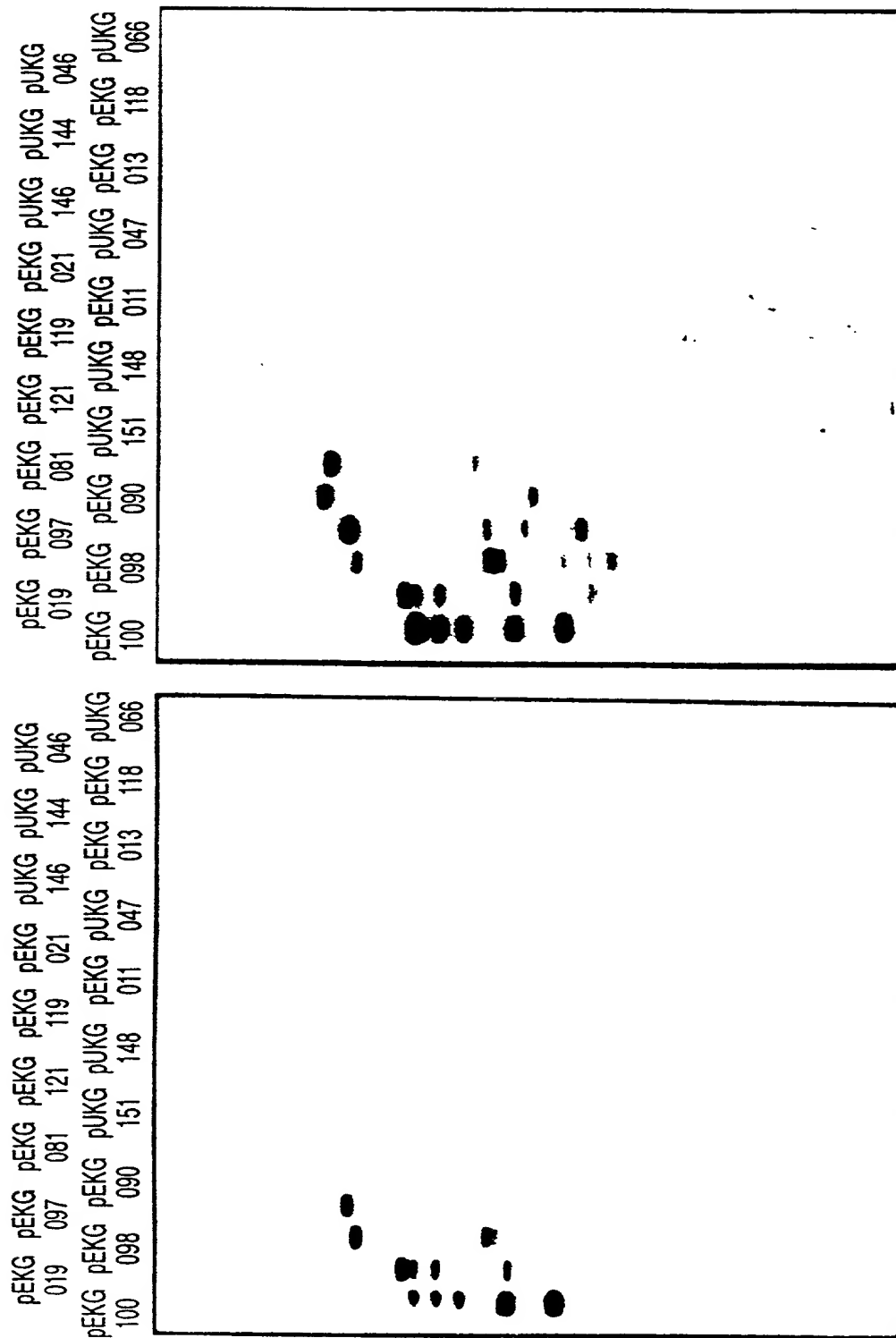


FIG. 17C

FIG. 17D

pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG  
 019 097 081 121 119 021 146 144 046 019 097 081 121 119 021 146 144 046  
 pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG pEKG  
 100 098 090 151 148 011 047 013 118 066 100 098 090 151 148 011 047 013 118 066

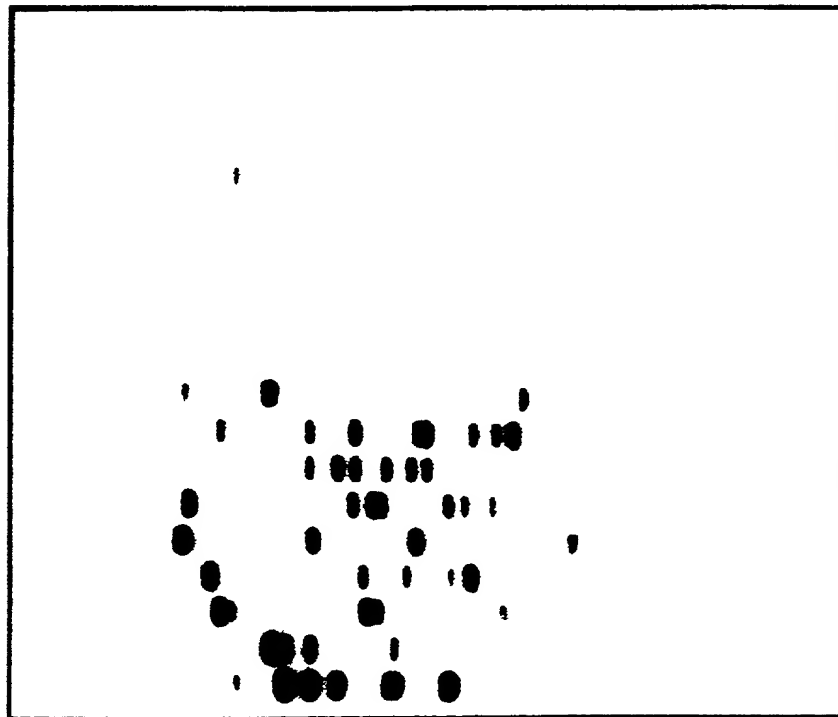


FIG. 17E

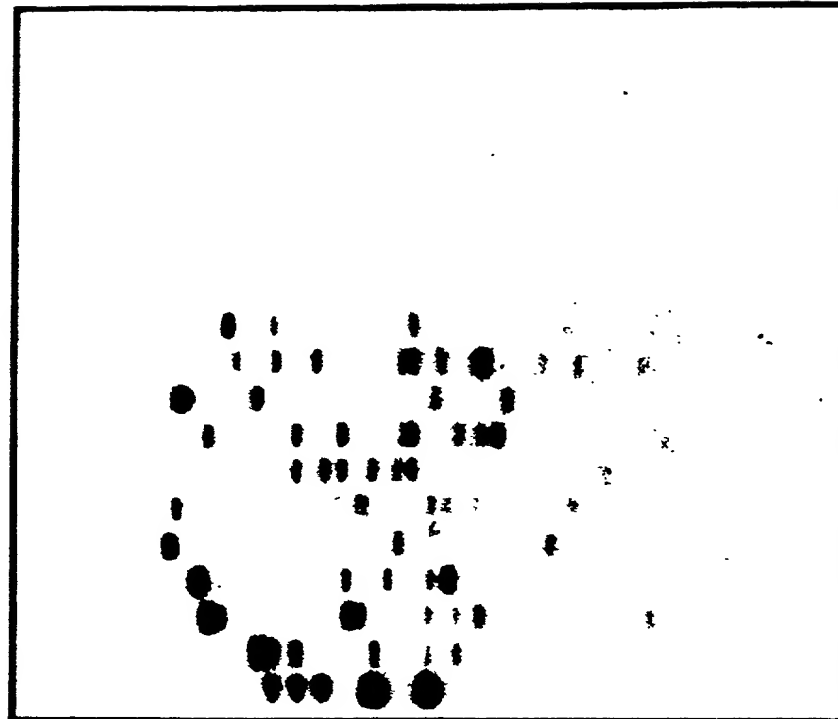


FIG. 17F

pEKG pEKG pEKG pEKG pEKG pEKG pUKG pUKG pUKG  
 019 097 081 121 119 021 146 144 046  
 pEKG pEKG pEKG pUKG pUKG pEKG pEKG pUKG  
 100 098 090 151 148 011 047 013 118 066

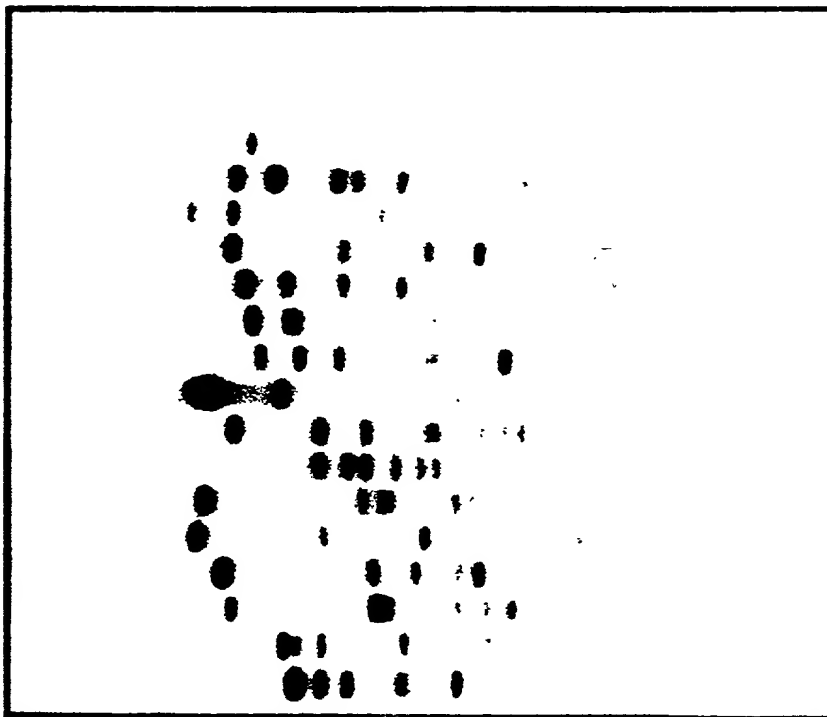


FIG. 17G

pEKG pEKG pEKG pEKG pEKG pEKG pUKG pUKG pUKG  
 019 097 081 121 119 021 146 144 046  
 pEKG pEKG pEKG pUKG pUKG pEKG pEKG pUKG pUKG  
 100 098 090 151 148 011 047 013 118 066

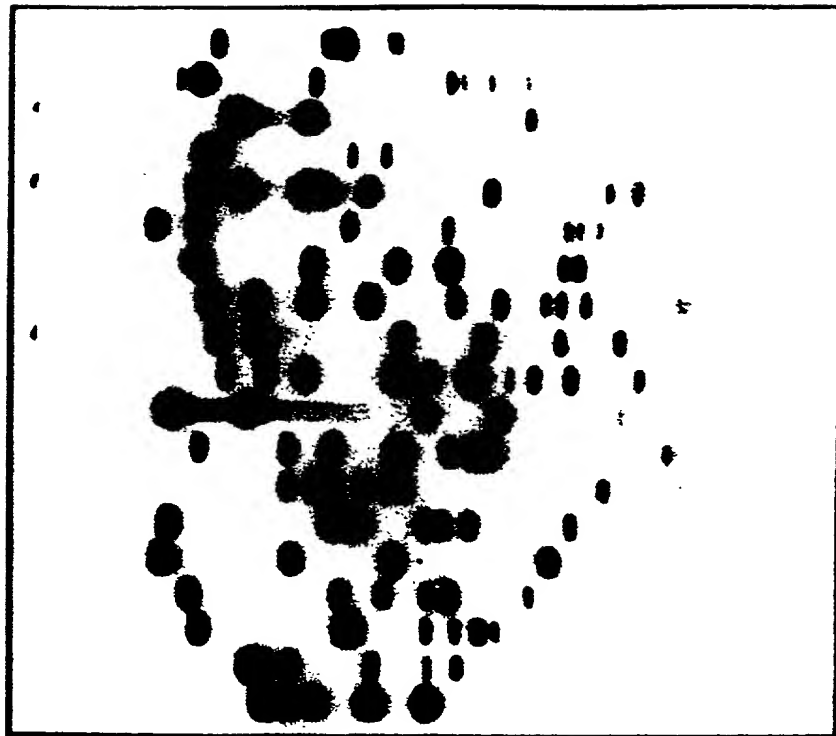


FIG. 17H

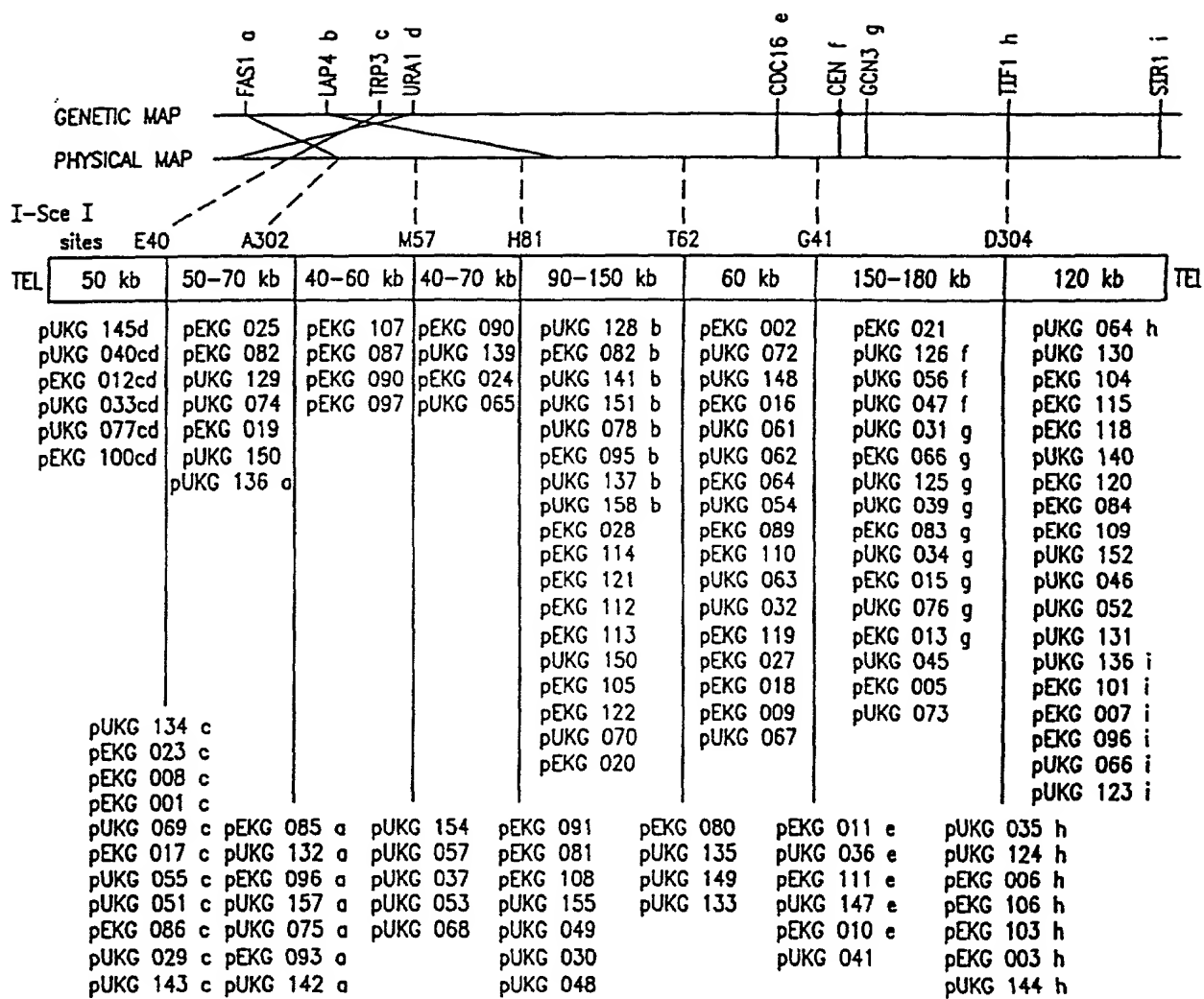


FIG. 18

